

UNCLASSIFIED

AD

401 480

*Reproduced
by the*

DEFENSE DOCUMENTATION CENTER

FOR

SCIENTIFIC AND TECHNICAL INFORMATION

CAMERON STATION, ALEXANDRIA, VIRGINIA



UNCLASSIFIED

NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

6332

CATALOGED BY ASTIA

480

4

1

4

0

1

4

0

1

4

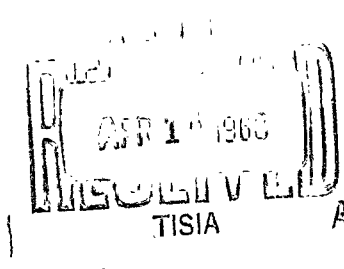
0

TECHNICAL

BULLETIN 62-15

**DEVELOPMENT AND STANDARDIZATION
OF FORM 6 OF THE U. S. NAVY
LITERACY TEST**

LEONARD SWANSON



AUGUST 1962



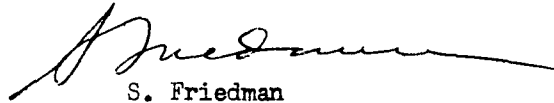
BUREAU OF NAVAL PERSONNEL

Submitted by

B. Rimland, PhD., Director, Personnel Measurement Research Department
E. E. Dudek, Ph.D., Technical Director
CDR H. B. Boaz, USN, Officer in Charge

U.S. Naval Personnel Research Activity
San Diego 52, California

Approved by



S. Friedman
Head, Psychological Research Branch



CAPT R. J. Baxter, USN
Director, Personnel Research Division

Bureau of Naval Personnel
Washington 25, D. C.

BRIEF

This report covers the development and standardization of a new form of the Literacy Test to replace the Literacy Test, Form X-1, (LIT X-1). Since its development in World War II, LIT X-1 has been used operationally as an aid in identifying Navy recruits with abilities too limited for retention in the Navy.

The new Literacy Test, Form 6, is a significant improvement over Form X-1 in terms of its difficulty level and discriminating power as well as in terms of its correlations with other tests of reading ability.

A new format incorporating reusable booklets and separate answer sheets was designed which will provide economy in maintaining test inventories and in scoring.

The Literacy Test, Form 6 was recommended for early operational use.

CONTENTS

	Page
A. STATEMENT OF THE PROBLEM	1
B. BACKGROUND	1
C. DEVELOPMENT OF THE TEST	
1. Guide for writing literacy items	2
2. Item revision and item writing	3
3. Literacy Test, Forms X-6A and X-6B	3
4. Administration of experimental Literacy Test forms .	4
5. The sample	4
6. Selection of items for Literacy Test, Form 6	4
D. STANDARDIZATION OF LITERACY TEST, FORM 6	5
E. TEST CHARACTERISTICS AND EVALUATION	
1. Item statistics	7
2. Correlations among tests	7
3. Means and standard deviations	8
4. Reliability	8
F. CONCLUSIONS AND RECOMMENDATIONS	11
REFERENCES	11
APPENDIX	

TABLES

	Page
1. Distributions of <u>p</u> -values and biserial correlations for items selected for LIT 6	6
2. Distributions of average <u>p</u> -values and item-total test correlations for LIT 6 items from the standardization administration	8
3. Intercorrelations and distribution statistics for literacy, reading and related tests for recruits to whom LIT 6 was administered before LIT X-1 (N=379) . .	9
4. Intercorrelations and distribution statistics for literacy, reading and related tests for recruits to whom LIT 6 was administered after LIT X-1 (N=390) . .	10
5. Literacy Test, Form 6 scores corresponding to Literacy Test, Form X-1 scores	A1
6. Literacy Test, Form 6 scores corresponding to Stanford Achievement Test, Intermediate Reading Test, Form K grade scores and grade equivalents	A2

THE DEVELOPMENT AND STANDARDIZATION OF FORM 6
OF THE U.S. NAVY LITERACY TEST¹

A. STATEMENT OF THE PROBLEM

The purpose of this project was to develop and standardize a new form of Literacy Test to replace Form X-1. This new form, using data from earlier forms as well as new items, was to have a difficulty level appropriate for evaluating the reading ability of incoming recruits who obtain a Navy Standard Score (NSS) of 35 or below on the General Classification Test (GCT).

B. BACKGROUND

The Literacy Test, Form X-1, (LIT X-1) is used operationally along with the Non-Verbal Classification Test (NVCT) and the U.S. Armed Forces Institute (USAFI) Reading Test for the early identification of Navy recruits with reading and/or mental ability too limited for retention in the Navy. Previously, the LIT X-1, together with the NVCT, had been used for selecting from among low ability recruits those who would receive literacy training in the Recruit Preparatory Training Program. This program is not in operation at the present time.

LIT X-1 was developed during World War II. It consists of 20 word recognition items, 16 sentence reading items and 8 paragraph comprehension items, all of which have four response alternatives. In the word recognition items, the examinee encircles the word representing the object pictured. In the sentence reading items, the examinee encircles the sentence that best tells what the picture shows. For the paragraph comprehension items, the examinee reads a paragraph and selects the group of words which tells best the main idea of the passage. Each part is separately timed. The score is the total number of items answered correctly. The test was standardized against the Gates Reading Survey Test (1939).

A revision, Form X-2, was begun toward the end of the war but was never completed. Later, in 1954, under an Office of Naval Research contract, the Institute of Educational Research, University of Indiana, developed two new Literacy Test Forms, X-3 and X-4 (Fattu and Standlee, 1954). These forms used items from Forms X-1 and X-2 as well as new material. In 1956, Forms X-1, X-3 and X-4 were given to an appropriate sample of Navy recruits

¹The early stages of this project, including item development and design of the new test format, were the work of Mrs. Vinnie H. Miller.

for analysis (Miller, 1957). The average item difficulty levels for LIT X-1, X-3 and X-4 ranged from .93 to .83. The standard deviations of the tests ranged from 2.8 to 4.1. Form X-1 has 44 items and Forms X-3 and X-4 each have 54 items.

Analysis of the available test items revealed that it would be possible to assemble a test form having a predicted mean item difficulty value of .63 and a predicted standard deviation of 5.3 (using no items with difficulty value over .90). Such a test was planned with the designation Literacy Test, Form X-5, but, upon further examination of the items available, it was felt that further work was necessary. Consequently Form X-5 was not administered, but served only as a base for the development of an improved form, designated Literacy Test, Form X-6.

C. DEVELOPMENT OF THE TEST

1. Guide for Writing Literacy Items

Some of the principles of achievement test construction must be re-examined when writing literacy items. In the usual achievement test, reading difficulty is held to a minimum except for necessary technical language. In a literacy test, however, word difficulty and sentence length and complexity become variables. In most achievement items it is undesirable to have extraneous clues to the correctness of an alternative. In literacy items, however, every word is a clue.

One index of word difficulty is the frequency of its occurrence per million words of general reading matter (Thorndike and Lorge, 1944). This index must be considered only as a rough guide, however, since a number of factors apart from frequency also influence word difficulty. For example, common nouns denoting everyday material objects are probably easier than their frequency would indicate. Generic nouns denoting a broad class of objects are probably more difficult than their specific counterparts. For example, "bicycle" is probably easier to read but occurs less frequently than "vehicle."

Polysyllabic words are in general more difficult than monosyllabic words, but words compounded of easy nouns are probably easier than their frequency would indicate. For example, "sailboat" is easier than "stipulate" although both have the same frequency.

The guide developed to aid in writing literacy items included the following considerations:

a. The items should vary, within prescribed limits, in word difficulty, sentence complexity and length.

b. The alternatives for a single item should be balanced in word difficulty, sentence complexity and length.

c. The correct alternative will be either the only true answer or the most likely and relevant alternative, its position to be assigned non-systematically.

d. Item and alternative content should be common or general knowledge and should be meaningful.

e. Incorrect alternatives should bear some relevance to the illustration, otherwise clue words might make them too easy to reject.

2. Item Revision and Item Writing

The item pool was reviewed to identify items meeting the above requirements. A sufficient number of paragraph comprehension items appeared to be satisfactory for inclusion in Form X-6 with only minor revisions. The sentence reading items, however, appeared to require considerable revision.

The finished test was to have from 15 to 20 sentence reading items and approximately 15 paragraph comprehension items. To allow for attrition in the sentence reading items, 18 items were revised and 18 new items were written for a total of 36 sentence reading items. Seventeen paragraph comprehension items were selected from Forms X-3 and X-4, a few of which were slightly revised, for inclusion in Form X-6. All items were in four alternative multiple-choice form.

3. Literacy Test, Forms X-6A and X-6B

To reduce the amount of testing time per subject, two experimental test booklets were prepared. One booklet contained 18 sentence reading items and the 17 paragraph comprehension items; the second contained the other 18 sentence reading items and the same set of 17 paragraph comprehension items. The booklets, each with two separately timed parts, were similar to the desired final form with respect to the number of each type of item, so that booklet format could be evaluated and reasonably good estimates of time requirements could be made.

Reusable booklets were designed which used a staggered page format and a separate removable answer sheet at the end of each part. Items were printed on the right hand page only. The first page of items stopped one-half inch short of the right hand edge of the booklet making the first column on the answer sheet visible. The next page of items was one-half inch narrower than the preceding page, revealing a second column on the answer sheet. Each succeeding page was similarly narrower than the preceding one to permit the use of a separate answer sheet.

These reusable test booklets reduce the cost and inconvenience of maintaining a sufficient inventory of test booklets. The use of a separate answer sheet reduces the time required for scoring by a significant extent.

4. Administration of Experimental Literacy Test Forms

LIT X-6A and LIT X-6B were administered from 26 August to 2 December 1957 to each recruit taking LIT X-1 at the Naval Training Center, San Diego. The two experimental Literacy Test forms were administered on alternate testing days following the administration of LIT X-1, the NVCT, and the USAFI Reading Test. Administering each experimental test form on alternate testing days assured that the recruit sample tested with each form was comparable and that after any given interval approximately an equal number of subjects would have taken each form. Sufficient time was allowed (15 minutes for Part I and 40 minutes for Part II) to permit everyone to finish the test.

5. The Sample

LIT X-6A and LIT X-6B were administered to 235 and 216 recruits, respectively. Data for 69 recruits were eliminated because they were Filipinos or Guamanians, because of missing data, or because the recruits were, in the opinion of the examiner, obviously intentionally trying to fail the test. Data for an additional 70 recruits scoring 36 or above on GCT were not included in the primary item analysis samples. The primary samples upon which item analyses were based were 161 recruits taking LIT X-6A and 151 recruits taking LIT X-6B, all of whom scored 35 or below on GCT and had complete test data available.

6. Selection of Items for Literacy Test, Form 6

In spite of its limitations, LIT X-1 has been considered to be a useful test for screening recruits with limited literacy abilities, and it was considered desirable to make use of it in evaluating the new items. Because of its limited difficulty level, however, LIT X-1 scores were combined with the USAFI Reading Test scores for evaluating LIT X-6 items. Since the score distributions of these two tests are quite different (the standard deviation of the USAFI Reading Test being much greater than that for the LIT X-1 Test), the raw scores for both tests were transformed to standard scores before combining. (Test security precluded the use of the USAFI Reading Test as a substitute for the LIT X-1 Test as a screening instrument for low level recruits.)

Biserial correlations were computed between the LIT X-6 items and the sum of standard scores on the LIT X-1 and USAFI Reading Tests for evaluating the items in the experimental forms.

Item selection was made on the basis of the item discrimination indices, item difficulty values, and on the distributions of responses among the alternatives.

Ten sentence reading items from each of the two experimental forms were selected for inclusion in the final form. None of these items had been included in Form X-1 of the Literacy Test. Fifteen of the 17 paragraph comprehension items in the experimental forms were selected for inclusion in the final form. Four of these items are very similar or identical to Form X-1 items.

Distributions of difficulty values and item biserial correlations for the selected items, based on experimental form data, are presented in Table 1, below. The p-values range from .46 to .85 with median values of .61 and .63 for the sentence reading and paragraph comprehension items, respectively.

The item biserial correlations range from .28 to .65 with a median of .40 for the sentence reading items and the range was from .23 to .96 with a median of .50 for the paragraph comprehension items. The higher correlations of the paragraph comprehension items are a reflection, in part, of overlapping items in the new test and the criterion. It should be noted that the sample involved was composed of Navy recruits scoring below 36 on the GCT.

D. STANDARDIZATION OF LITERACY TEST, FORM 6

The standardization of LIT 6 was based on a sample of about 800 recruits, 400 each from NTC/Great Lakes and NTC/San Diego, scoring 35 or below on GCT, who were tested from October 1960 to January 1961. In order to balance any practice effects which might occur from taking one test before the other, LIT 6 was administered before LIT X-1 to approximately one-half of the examinees at each NTC (N=390). The time limit of 15 minutes for Part I and 35 minutes for Part II permitted virtually everyone to finish the test.

The Stanford Reading Test, Intermediate Reading Test, Form K, was administered following both Literacy Test forms to permit relating LIT 6 scores to grade level scores.

The conversion table was constructed by the equi-percentile method. Separate graphs were drawn for each sample to examine the effects of practice on a similar test. There appeared to be no gain in raw score attributable to prior practice on any of the reading tests.

LIT 6 scores corresponding to LIT X-1 scores and to Stanford Reading grade scores and to grade equivalents are presented in Tables 5 and 6 in the appendix.

TABLE 1
Distributions of p-Values and Biserial Correlations
for Items Selected for LIT 6

<u>p</u> -Value or Fit value	Sentence Reading Items ^a		Paragraph Comprehension Items ^{b,c}	
	Distribution of <u>p</u> -values	Distribution of <u>r</u> _{bis}	Distribution of <u>p</u> -values	Distribution of <u>r</u> _{bis}
.90 and above				1
.80 - .89	2		2	
.70 - .79	3		1	3
.60 - .69	7	1	8	1
.50 - .59	3	4	2	3
.40 - .49	5	5	2	2
.30 - .39		7		3
.20 - .29		3		2
No. of items	20		15	
Median value ^d	.61	.40	.63	.50

^aData for Sentence Reading items were based on an N of 161 or an N of 151.

^bData for the Paragraph Comprehension items were based on an N of 312. Biserial correlations and p-values were averaged for the two samples of 161 and 151. Biserial correlations were uncorrected for inclusion of items in LIT X-1.

^cFour of the LIT 6 Paragraph Comprehension items are identical or very similar to LIT X-1 items.

^dMedians were computed from raw data.

E. TEST CHARACTERISTICS AND EVALUATION

1. Item Statistics

In order to evaluate the final test at the item level, two samples ($N = 370$ each) were selected from the standardization sample. One sample had taken LIT 6 before LIT X-1 and the other LIT 6 after LIT X-1. From each of these samples the high and low 27 per cent of each sample were selected. Item difficulty values (based on the entire group of 370) and item-total test correlation coefficients (Flanagan r 's) were determined for each item. Because the statistics obtained on the two samples were very similar, the values obtained on the two samples were averaged. Distributions of the obtained p -values and r_{it} 's are presented in Table 2 below. The estimated item difficulty values ranged from .36 to .94 with a median of .71. The r_{it} 's ranged from .21 to .65 with a median of .50.

Statistics for the two parts of the test indicate that the paragraph reading items are a little easier (average p -value of .74 as contrasted to .63 for sentence reading items) and have slightly higher item-test correlations (median $r_{it} = .54$) than the sentence reading items (median $r_{it} = .48$).

2. Correlations Among Tests

Intercorrelations and distribution statistics for LIT 6, Stanford Reading and operational tests were obtained on the two parts of the standardization sample (N 's = 379 and 390) to assist in the evaluation of LIT 6.

Variables included in the matrices were GCT, NVCT, LIT X-1, USAFI Reading, LIT 6 part and total scores, and Stanford Reading part and average scores. These data are presented in Tables 3 and 4, below.

The correlations between LIT X-1 and LIT 6 were .50 and .44 in the two samples. Since the two tests differ substantially in difficulty these should not be thought of as alternate-form reliabilities.

Of particular interest are the correlations between LIT tests and the Stanford Reading Test. LIT X-1 scores and the Stanford Reading average scores correlate .41 and .48 in the two samples, while LIT 6 scores and the Stanford Reading average scores correlate .76 and .73. This very substantial improvement in the correlation of the Literacy Test with the widely accepted Stanford Reading Test is highly significant statistically and may properly be interpreted as an indication of an increase in validity of Form 6 over Form X-1.

Correlations between LIT scores and GCT or NVCT scores are quite similar for both LIT Tests.

TABLE 2

Distributions of Average \bar{p} -Values and Item-Total Test Correlations^a
 For LIT 6 Items From the Standardization Administration
 (N=740)

\bar{p} -value or \bar{r}_{it} value	Sentence Reading Items		Paragraph Compre- hension Items		Total Test	
	Distribution of		Distribution of		Distribution of	
	\bar{p} -values	\bar{r}_{it} 's	\bar{p} -values	\bar{r}_{it} 's	\bar{p} -values	\bar{r}_{it} 's
.90 - .99	1		1		2	
.80 - .89	5		2		7	
.70 - .79	3		7		10	
.60 - .69	3	3	2	4	5	7
.50 - .59	4	5	2	6	6	11
.40 - .49	2	9	1	3	3	12
.30 - .39	2	1		2	2	3
.20 - .29		2				2
No. of items	20		15		35	
Median value ^b	.64	.48	.74	.54	.71	.50

^aItem-total test correlations are averages of Flanagan \bar{r} 's based on high and low 100 of each of the two samples of 370.

^bMedians were computed from raw data.

3. Means and Standard Deviations

The 35-item LIT 6 test is a far more discriminating test than is the 44-item LIT X-1 test, as is indicated by the relative standard deviations of the tests. For LIT X-1, the standard deviations in the two samples are 2.93 and 2.47. Corresponding standard deviations for LIT 6 are 5.39 and 5.82. The means of about 41 and 24 for LIT X-1 and LIT 6 represent average item difficulties of .93 and .68, respectively.

4. Reliability

Internal consistency reliability estimates were computed from data obtained on the two samples of 370 used in item analysis. Kuder-Richardson reliabilities for LIT 6, using Formula 21, were .80 for one sample and .76 for the other. While these values are

TABLE 3

Intercorrelations and Distribution Statistics for Literacy, Reading and Related Tests for Recruits to Whom LIT 6 Was Administered Before LIT X-1 (N=379)

	GCT	NVCT	LIT X-1	USAFI Read.	LIT 6		LIT 6 Total	Stanford Reading		Means	Standard Deviations
					Part 1	Part 2		Part 1	Part 2		
GCT		20	09	15	12	13	13	18	18	32.37	2.80
NVCT	20		40	19	29	31	34	34	21	52.38	8.75
LIT X-1	09	40		28	39	40	44	37	39	41.18	3.47
USAFI Reading	15	19	28		41	42	46	46	47	76.32	27.43
LIT 6 Part 1	12	29	39	41		63	92	59	72	13.36	3.56
LIT 6 Part 2	13	31	40	42	63		87	57	62	10.82	2.83
LIT 6 Total	13	34	44	46	92	87		64	75	24.13	5.82
Stanford Read. Part 1	18	34	37	46	59	57	64		68	52.77	11.60
Stanford Read. Part 2	18	21	39	47	72	62	75	68		62.20	13.79
Stanford Read. Average	20	30	41	52	72	65	76	90	93	57.40	11.77

NOTES--

1. Decimal points have been omitted from correlations.
2. The full matrix is presented to facilitate use of the table.

TABLE 4

Intercorrelations and Distribution Statistics for Literacy, Reading and Related Tests for Recruits to Whom LIT 6 Was Administered After LIT X-1 (N=390)

	GCT	NVCT	LIT X-1	USAFI Read.	LIT 6 Part 1	LIT 6 Part 2	LIT 6 Total	Stanford Reading		Means	Standard Deviations
								Part 1	Part 2		
GCT		14	13	08	12	03	09	13	21	19	3.28
NVCT	14		23	19	16	22	21	33	18	27	8.92
LIT X-1	13	23		33	48	42	50	41	47	48	2.93
USAFI Reading	08	19	33		36	34	39	42	43	48	26.45
LIT 6 Part 1	12	16	48	36		58	91	56	65	66	3.36
LIT 6 Part 2	03	22	42	34	58		86	54	62	64	2.70
LIT 6 Total	09	21	50	39	91	86		62	72	73	5.39
Stanford Read.											
Part 1	13	33	41	42	56	54	62		66	90	10.86
Stanford Read.											
Part 2	21	18	47	43	65	62	72	66		92	12.53
Stanford Read.											
Average	19	27	48	48	66	64	73	90	92		10.74

NOTES--

1. Decimal points have been omitted from correlations.
2. The full matrix is presented to facilitate use of the table.

somewhat lower than are desired, Kuder-Richardson estimates are considered a lower bound of reliability.

Reliability estimates for LIT X-1 were .80 and .72 for the corresponding samples.

F. CONCLUSIONS AND RECOMMENDATIONS

LIT 6 is a significant improvement over LIT X-1 and should be substituted for it as soon as feasible for the following reasons:

a. The correlation between LIT 6 and the Stanford Reading Test (average $r = .76$) is significantly higher than that between LIT X-1 and the Stanford Reading Test (average $r = .45$); therefore, LIT 6 can be considered a more satisfactory measure of reading achievement than Form X-1.

b. The standard deviation has been increased from about 3.3 in LIT X-1 to about 5.5 in LIT 6.

c. The items in LIT 6 are of a more appropriate difficulty level than those in LIT X-1 for use with recruits with GCT below a Navy Standard Score of 36, the average difficulty values being .68 and .93, respectively.

d. The internal consistency estimates of reliability appear to be satisfactory for LIT 6.

e. Savings in scoring time and in material will be effected through the replacement of LIT X-1 by LIT 6.

REFERENCES

- Fattu, N. A. & Standlee, L. S. Rationale and procedure for developing a Navy literacy test: illustrated by development of two parallel test forms. Institute of Educational Research, Indiana University; prepared under Contract Nonr-908(01), July 1954.
- Gates, A. I. Gates Reading Survey: Manual. New York: Columbia University Press, 1939.
- Miller, Vinnie H. An evaluation of the Navy Literacy Test, Forms X-3 and X-4. U. S. Naval Personnel Research Field Activity, Report No. 108, January 1957.
- Thorndike, E. L. & Lorge, I. Teachers' Word Book of 30,000 Words. Teachers College, Columbia University, 1944.

APPENDIX

TABLE 5

Literacy Test, Form 6, Scores Corresponding
To Literacy Test, Form X-1 Scores

LIT 6 Score	LIT X-1 Score
32-35	44
29-31	43
25-28	42
22-24	41
19-21	40
16-18	39
15	38
14	37
13	36
12	35
11	33
10 and below	28

APPENDIX

TABLE 6

Literacy Test, Form 6 Scores Corresponding to Stanford
Achievement Test, Intermediate Reading Test, Form K
Grade Scores and Grade Equivalents

LIT 6 Score	Stanford Reading Test		
	Grade Score	Grade Equivalents	
		Par. Mean. (Part 1)	Word Mean. (Part 2)
35	98	10.1	10.3
34	90	9.5	9.5
33	82	8.7	8.8
32	78	8.3	8.3
31	73	7.8	7.8
30	70	7.5	7.5
29	68	7.3	7.3
28	65	7.0	6.9
27	63	6.7	6.7
26	61	6.5	6.5
25	59	6.2	6.3
24	57	6.0	6.1
23	55	5.8	5.9
22	53	5.6	5.6
21	51	5.4	5.4
20	49	5.1	5.2
19	47	4.9	5.0
18	46	4.8	4.9
17	44	4.6	4.7
16	43	4.5	4.6
15	42	4.4	4.2
14	40	4.2	4.2
13	38	4.0	4.0
12	36	3.8	3.8
11	33	3.5	3.5
10	31	3.3	3.3
9 and below	29	3.1	3.1